Class11 HW

Isabella Ruud PID A59016138

Q13: Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

Hint: The read.table(), summary() and boxplot() functions will likely be useful here. There is an example R script online to be used ONLY if you are struggling in vein. Note that you can find the medium value from saving the output of the boxplot() function to an R object and examining this object. There is also the medium() and summary() function that you can use to check your understanding.

```
#read in txt file
expr <- read.table("sample geno exp.txt")

#look at file to make sure it looks okay
head(expr)
```

```
   sample geno      exp
1 HG00367  A/G 28.96038
2 NA20768  A/G 20.24449
3 HG00361  A/A 31.32628
4 HG00135  A/A 34.11169
5 NA18870  G/G 18.25141
6 NA11993  A/A 32.89721
```

```
#get the sample size for each genotype
table(expr$geno)
```

```
A/A A/G G/G
108 233 121
```

There are 108 A/A genotypes, 233 A/G genotypes, and 121 G/G genotypes.

```
#get the median expression values for each genotype
aa_median <- median(expr[expr$geno == "A/A",]$exp)
ag_median <- median(expr[expr$geno == "A/G",]$exp)
gg_median <- median(expr[expr$geno == "G/G",]$exp)

cat("A/A expression median: ", aa_median, "\n")
```

A/A expression median: 31.24847

```
cat("A/G expression median: ", ag_median, "\n")
```

A/G expression median: 25.06486

```
cat("G/G expression median: ", gg_median, "\n")
```

G/G expression median: 20.07363

A/A expression median: 31.24847 A/G expression median: 25.06486 G/G expression median: 20.07363

```
#can also get the summary stats (mean, median, quartiles, etc)
cat("A/A genotype:", "\n")
```

A/A genotype:

```
summary(expr[expr$geno == "A/A",])
```

```
    sample              geno                  exp
 Length:108         Length:108         Min.   :11.40
 Class :character   Class :character   1st Qu.:27.02
 Mode  :character   Mode  :character   Median :31.25
                                       Mean   :31.82
                                       3rd Qu.:35.92
                                       Max.   :51.52
```

```
cat("A/G genotype:", "\n")
```

A/G genotype:

```
summary(expr[expr$geno == "A/G",])
```

```
    sample              geno                  exp
 Length:233         Length:233         Min.   : 7.075
 Class :character   Class :character   1st Qu.:20.626
 Mode  :character   Mode  :character   Median :25.065
                                       Mean   :25.397
                                       3rd Qu.:30.552
                                       Max.   :48.034
```

```
cat("G/G genotype:", "\n")
```

```
G/G genotype:
```

```
summary(expr[expr$geno == "G/G",])
```

```
    sample              geno                  exp
 Length:121         Length:121         Min.   : 6.675
 Class :character   Class :character   1st Qu.:16.903
 Mode  :character   Mode  :character   Median :20.074
                                       Mean   :20.594
                                       3rd Qu.:24.457
                                       Max.   :33.956
```

Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORMDL3?

Hint: An example boxplot is provided overleaf – yours does not need to be as polished as this one.

```
library("ggplot2")
```

```
#make a box plot of the ORDML3 expression data for each genotype

ggplot(expr) +
  aes(x = geno, y = exp, fill = geno) +
  geom_boxplot(notch = TRUE) +
  geom_jitter(width = 0.2, height = 0, alpha = 0.25) +
  labs(x="Genotype", y = "Expression", title = "SNP Effect on ORDML3 Expression", fill = "Ger
  theme(plot.title = element_text(hjust = 0.5))
```

SNP Effect on ORDML3 Expression
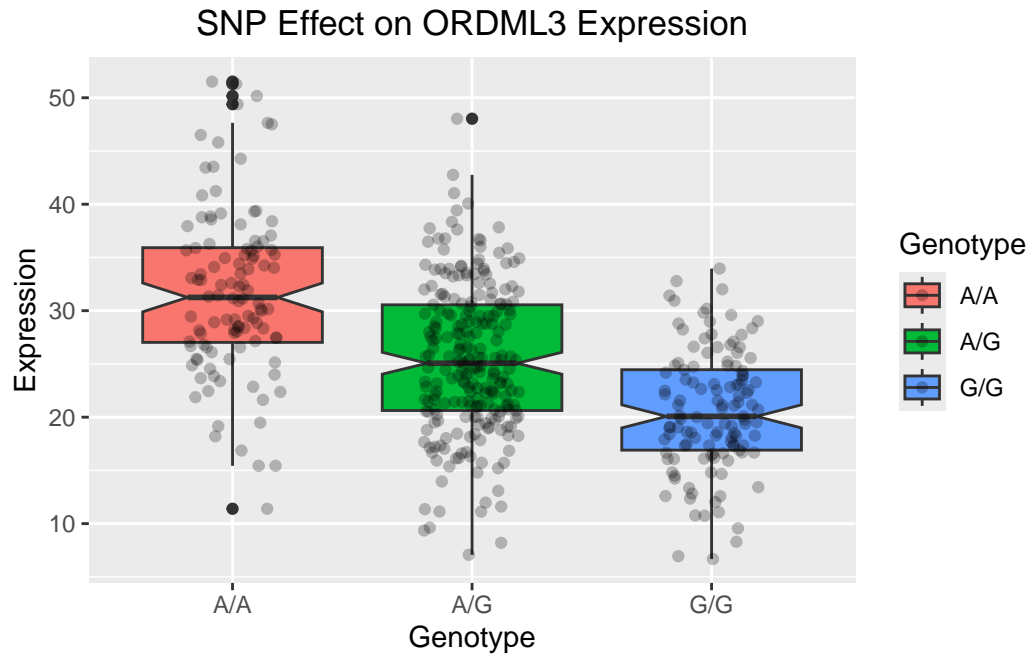
Yes, the SNP has an effect on the relative expression of ORDML3. The A/A genotype has the highest expression and the G/G genotype has the lowest expression with A/G sitting in between them. It seems that as the number of G alleles increases, the expression of ORDML3 decreases.